

Conference Abstract

Improving Impact Metrics of Open and Free Biodiversity Data through Linked Metadata and Academic Outreach

Daniel Noesgaard [‡]

[‡] GBIF Secretariat, Copenhagen, Denmark

Corresponding author: Daniel Noesgaard (dnoesgaard@gbif.org)

Received: 24 Apr 2019 | Published: 13 Jun 2019

Citation: Noesgaard D (2019) Improving Impact Metrics of Open and Free Biodiversity Data through Linked Metadata and Academic Outreach. Biodiversity Information Science and Standards 3: e35723.

<https://doi.org/10.3897/biss.3.35723>

Abstract

The work required to collect, clean and publish biodiversity datasets is significant, and those who do it deserve recognition for their efforts. Researchers publish studies using open biodiversity data available from GBIF—the Global Biodiversity Information Facility—at a rate of about two papers a day. These studies cover areas such as macroecology, evolution, climate change, and invasive alien species, relying on data sharing by hundreds of publishing institutions and the curatorial work of thousands of individual contributors. With more than 90 per cent of these datasets licensed under Creative Commons Attribution licenses (CC BY and CC BY-NC), data users are required to credit the dataset providers. For GBIF, it is crucial to link these scientific uses to the underlying data as one means of demonstrating the value and impact of open science, while seeking to ensure attribution of individual, organizational and national contributions to the global pool of open data about biodiversity.

Every single authenticated download of occurrence records from GBIF.org is issued a unique Digital Object Identifier (DOI). These DOIs each resolve to a landing page that contains details of

- the search parameters used to generate the download

- a quantitative map of the underlying datasets that contributed to the download
- a simple citation to be included in works that rely on the data

When used properly by authors and deposited correctly by journals in the article metadata, the DOI citation establishes a direct link between a scientific paper and the underlying data. Crossref—the main DOI Registration Agency for academic literature—exposes such links in Event Data, which can be consumed programmatically to report direct use of individual datasets. GBIF also records these links, permanently preserving the download archives while exposing a citation count on download landing pages that is also summarized on the landing pages of each contributing datasets and publishers. The citation counts can be expanded to produce lists of all papers unambiguously linked to use of specific datasets.

In 2018, just 15 per cent of papers based on GBIF-mediated data used DOIs to cite or acknowledge the datasets used in the studies. To promote crediting of data publishers and digital recognition of data sharing, the GBIF Secretariat has been reaching out systematically to authors and publishers since April 2018 whenever a paper fails to include a proper data citation. While publishing lags may hinder immediate effects, preliminary findings suggest that uptake is improving—as the number of papers with DOI data citations during the first part of 2019 is up more than 60 per cent compared to 2018.

Focusing on the value of linking scientific publications and data, this presentation will explore the potential for establishing automatic linkage through DOI metadata while demonstrating efforts to improve metrics of data use and attribution of data providers through outreach campaigns to authors and journal publishers.

Keywords

data citation, attribution, credit, GBIF, biodiversity data article metadata

Presenting author

Daniel Noesgaard